The Role of Confidence in T&E of Strategic Defense Systems

Larry J. Levy

The Johns Hopkins University
Applied Physics Laboratory
Laurel, Maryland 20723

Abstract

Weapon system confidence is being able to predict the system's performance to within a quantified uncertainty (confidence interval). Properly planned test and evaluation of the system allows for models and simulations to be built to predict system performance with confidence. As confidence is important to strategic offensive weapons, it is equally important for defense against strategic warheads. Steps for building in confidence start with specifying the top-level family of systems performance evaluation requirements in terms of confidence. These are then "flowed down" to lower level system/subsystem performance requirements (confidence) using force on force level simulations. Test programs (test size, instrumentation quality) and analysis methodologies are then designed to meet each lower level requirement. Such a process provides for tradeoffs to be made while quantifying the implications of decisions to test more, or less, to instrument different functions or systems, or to changing the quality of the instrumentation. The fundamental feature of this test and evaluation process is to build models with associated confidence for the family of systems from which credible performance predictions can be made with quantified confidence intervals. This will allow for optimum planning for placement and usage of assets before the action commences as well as optimum real-time threat response. However, a number of "grand" technical challenges must be faced in order to optimally build in confidence to the ballistic missile defense family of systems.

Introduction

The national command recognized that not understanding how well our strategic deterrent (offensive) systems would perform (i.e., with quantified confidence) would be unacceptable. Therefore they set specific guidelines for test and evaluation of these systems, IDA/WSEG (1966)[1]. Analogous guidelines are <u>not</u> generally applied to tactical systems. The consequences of not knowing how well our strategic <u>defensive</u> systems will perform are at least as disastrous as they would have been for our deterrent systems**.** Therefore equivalent or more comprehensive guidelines should be promulgated for defense against strategic warheads. We need to <u>credibly</u> predict the operational performance of our deployed BMD systems. This is not just "how well will they will perform?" but "how <u>confident</u> are we in our prediction?"

Quantified confidence in performance assessments played a significant role in the development, testing and maintenance of the Trident II Weapon System; it should play an even more critical role for high value strategic defense systems such as Theatre Missile Defense (TMD) and National Missile Defense (NMD). Quantified confidence is knowing the system's performance to within a quantified uncertainty (confidence interval). It is statistically knowing what you don't know about the system performance. Building a weapon system with a good performance estimate (e.g. high reliability) but with a large confidence interval (high uncertainty) about that estimate could be disastrous!

Our missile defense systems must protect our troops and/or homeland against nuclear and/or biochemical warhead missiles (e.g., from major national conflicts, rogue nation, or terrorist attacks). Our systems must work the first time! The US public will not allow for any disasters. We must prevent the threat from holding the US public and government hostage in peace negotiations (This could happen if we suspect our system is not as good as planned.). Military planners and US policy makers need quantified confidence in the weapon systems performance estimates. It focuses attention to critical problems/subsystems where more testing could be applied. It provides the necessary information for optimization in the use of weapon assets for real-time response to a threat or for defense planning. High confidence provides high assurance for policy negotiations and high deterrence to potential adversaries. The question is not, "Can we afford to build in confidence?" but, "Can we afford <u>not</u> to

[1] These guidelines have evolved; the most recent version is US Strategic Command (1998).

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
27MAR2001	N/A	27MAR2001 - 29MAR2001

Title and Subtitle The Role of Confidence in T&E of Strategic Defense Systems	**Contract Number**
	Grant Number
	Program Element Number
Author(s) Levy, Larry J.	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) The Johns Hopkins University Applied Physics Laboratory Laurel, MD 20723	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) OSD Pentagon Washington, DC	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from the Proceedings AIAA 2nd Biennial National Forum on Weapon System Effectiveness, held at the John Hopkins University/Applied Physics Laboratory, 27-29 March 2001. Controlling Agency is OSD, Pentagon, Washington DC, per Richard Keith, Editor. See also ADM201408, non-print version (whole conference). , The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** SAR

Number of Pages
6

build confidence?" The cost increment to accomplish this is minimal.

Fundamental concepts of confidence in ballistic missile defense (BMD) system performance prediction are presented. Then an example in TMD will be summarized to illustrate the need for confidence based evaluation and prediction. Next, the actual application of confidence based methods to the test and evaluation of the Trident II Weapon System will demonstrate how it has been successfully done. Finally, an outline of a proposed approach for missile defense coupled with a discussion of the new technical challenges this presents is given.

Confidence in BMD Performance Prediction

The top level Measure of Effectiveness (MOE) for BMD is probability of negation, P_n, (or protection effectiveness). It is related to lower level system and subsystem parameters such as accuracy, reliability, time delays and many others. These are statistical parameters that are conceptually based on identically repeated trials of the family of systems (FoS) campaign scenario. As the number of trials gets large, estimates of these parameters converge to their true fixed values. For example, the estimator, $\hat{P}_n$, which equals the ratio of number of successful kills to the total number of threats, converges to the true underlying P_n. The parameters are "fixed but unknown" but our estimates of these parameters based on limited trials (testing) will be stochastic in nature with their distributions, being defined by the estimator forms, the quality of the instrumentation and the number of trials. The estimator distributions describe our uncertainty about the truth with confidence bounds (or intervals) being specific expressions of that uncertainty. Since some of the higher-level MOEs are not practically testable, e.g. many-on-many FoS performance at the theatre level, system evaluators must use simulation models to project FoS performance from lower level testable parameter distributions to the theatre level.

An example of the estimation distribution for P_n is shown in Figure 1, which is a binomial density function with its maximum near the true value. $\hat{P}_n$ is estimated from a specific test program. Conceptually, if the test program is repeated many times, randomness in the system and instrumentation would produce the distribution of $\hat{P}_n$ about the fixed value of P_n. As the number of tests in the test program increases, the $\hat{P}_n$ distribution gets more concentrated about the true P_n. An evaluation requirement might be to choose a test program such that $P_n - 90LCB < 0.1$. In reality, since the P_n for a specific campaign scenario is not directly testable, model simulations must be used to project $\hat{P}_n$ from lower level MOE estimates and their distributions. These may not necessarily look like the sampling density of Figure 1.

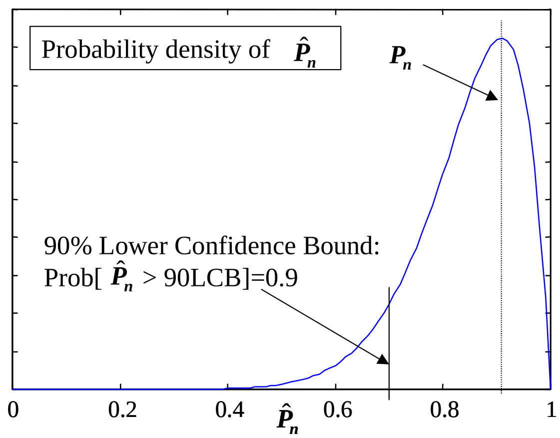

Figure 1 – Estimation Distribution for P_n

The overall concept of test program derived performance estimates and associated estimated distributions, allowing confidences to be estimated, is shown in Figure 2. Essentially, the sources of data provide modeling information to the system evaluator who constructs performance (e.g. accuracy, reliability

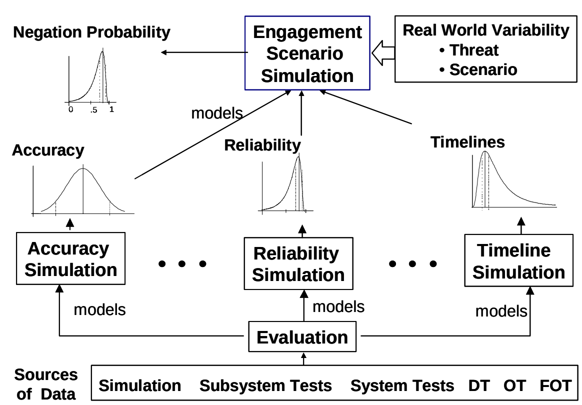

Figure 2 - BMD MOE Evaluation with Confidence

and timeline) estimates and associated distribution estimates. The test derived estimates and distributions are sampled for a Monte Carlo engagement scenario simulation, which transforms them into the top level MOE estimate, $\hat{P}_n$, and associated distribution. The construction of the performance estimates from the multiple sources of test data is not straight forward, requiring advanced system modeling techniques.

Importance of Confidence in a TMD Example

A companion paper in this conference, Mitchell et.al. (2001), provides an example of the importance of confidence in TMD. Starting with assumed estimates (called expected values in the paper) and associated distributions for accuracy, reliability and timelines, the Extended Air Defense Simulation (EADSIM) was used in a Monte Carlo mode, as illustrated in Figure 2, to evaluate a fictitious theatre engagement scenario. It was found that the variations in the sampled distributions could sometimes cause the FoS to perform radically different than predicted by just the projection of the assumed estimates. In other words, the reasonably possible statistical variations in the associated lower level distributions caused significant tails (and even multiple modes) in the $\hat{P}_n$ distribution. So it is possible that one could have predicted a reasonably high $\hat{P}_n$ with the true value being significantly lower, a potentially disastrous result! Again, <u>a properly constructed test program must be developed so as to achieve sufficiently close confidence bounds to the truth.</u>

Confidence in Trident II Accuracy Prediction

Goals for Trident II accuracy evaluation were specified in IDA/WSEG (1966) and the evaluation requirements were specifically defined in US Strategic Command (1998). The requirements specified quantified confidence goals for top-level MOE estimates of reliability and accuracy for initial performance estimates and change detection with time. For brevity, only accuracy evaluation will be described.

The process followed very closely the steps outlined in the next section except it was applied to the MOE of target accuracy. An overview description is given in Simkins et.al. (1990). New evaluation methodology (a satellite missile tracking system and maximum likelihood system identification for modeling) was developed to minimize system tests with greater functionality. Thirty system tests were needed using the traditional ("shoot and score") evaluation approach with only ten tests needed with the new methodology for initial model estimation. Ten tests were needed using traditional evaluation to four tests using the new methodology for detection of model changes in follow-on testing. Only the new methodology enabled extrapolation to untested conditions. Individual guidance error models and launch area gravity models were corrected. Increased system understanding was obtained to accurately predict performance over long-range non-tested trajectories. The estimated Trident II performance was considerably different than was expected. This would not have been known or understood with the traditional approach. This has enabled test-based predictions of capability to support other (non-traditional) missions & requirements.

Conceptual Application to T&E of BMD

The systems engineering approach to test and evaluation of BMD with confidence is shown in Figure 3. This was extrapolated from experience with many previous weapons systems T&E and especially that of Trident II. The left side illustrates the planning steps required to properly design an overall test program to provide adequate prediction confidence at certain milestones in the test program.

The key starting point is specifying the top level Performance Evaluation Requirements (not how well the weapon system should perform, but how well should we know it) in terms of required specifications (e.g. negation probabilities for realistic overall force level scenarios). At present, <u>there does not appear to be "official" evaluation requirements on how well we must know</u> P_n as there is for Trident II accuracy and reliability. This will be a serious impediment to successful employment of the BMD system. <u>A few test successes does not guarantee that the system will meet its objectives; it only shows that success is possible.</u>

If there is no top level MOE evaluation requirement in terms of confidence, then one must be developed. This would be an iterative process between developer, evaluator, and the military user. Questions to answer would be: What are the "required" performance values (e.g. negation probabilities) for realistic overall force level scenarios? How well do we need to know them? (i.e. width of the 90% confidence bounds?).

The next step is to determine a complete set of lower level Measures of Performance (MOPs) with associated confidence requirements over a reference

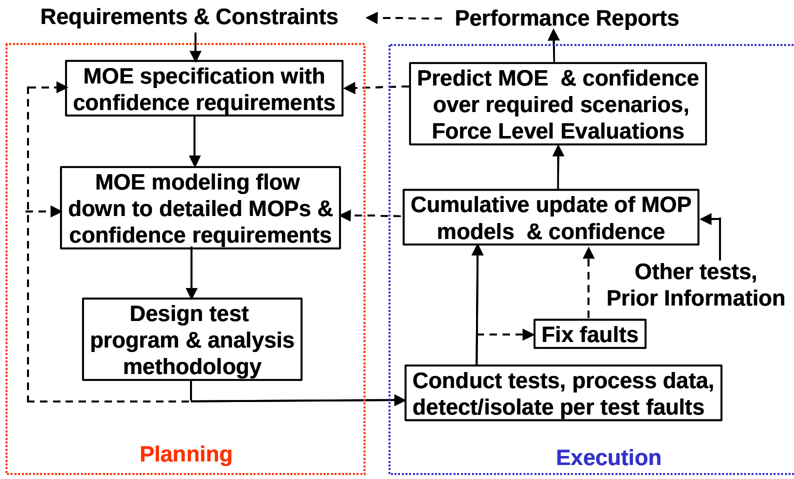

Figure 3 - Conceptual Approach to Test and Evaluation with Confidence

set of force level scenarios needed to achieve the required MOE & confidence bound. Testable MOPs (or ones that are extrapolated from tests) are sampled and force level simulations are used to flow up the MOPs (and confidences bounds) to the MOE (and confidence bounds). This process is iterated until an optimized set of MOPs (and confidence bounds) is achieved. The optimization criteria might be to "balance" the contributions of each MOP confidence contribution to MOE confidence. Other criteria might reflect the difficulty (e.g. cost) in achieving certain MOP confidence such as reliability. Many tradeoffs could be evaluated.

A test program and analysis methodology is then designed to meet each MOP confidence requirement by hypothesizing various feasible tests (system, sub-system, component), test sizes, instrumentation quality, and evaluation methodologies. Appropriate simulation models (covariance or Monte Carlo) are used to evaluate each hypothesized set until an optimized set is obtained. The results of this phase might require going back to the previous phase to revise the required MOP confidence bounds.

Such a process provides for tradeoffs to be made while quantifying the implications of decisions to test more, or less, to instrument different functions or systems, or to changing the quality of the instruments. As defense spending and costs associated with system development, test and evaluation come under increasing scrutiny, it becomes even more important to be able to quantify the relative benefits of test size and instrumentation quality. Quantifying the confidence with which we will know system performance provides a metric by which we can assess the value of our test programs, instrumentation and analysis approaches.

The right hand side of Figure 3 describes the execution steps in the test and evaluation process. Tests could be conducted by traditional testers and evaluators, but with the evaluation outputs complying with the system evaluator's requirements. Test types could include system, components or subsystem tests, monitoring of an in-place system as it awaits operational usage, and subsystems tested in-the-loop of a simulation. Per test fault detection/isolation would be conducted by traditional tester/evaluators, but with results validated by the system evaluator. Isolated faults would be fixed by the developer and removed from the data base and models.

The system evaluator would calculate a cumulative update of the MOP models, confidence intervals and estimated distributions. Use of physics based models, where possible, to fit data (system identification) from diverse tests would be used to gain maximum information from each test. If the model can be broken down to a set of parameters that are independent of scenario, then statistical leverage can be gained by accumulating across all relevant but disparate tests. This process for accuracy is described in Levy (1996). The associated uncertainty (confidence bound) in the model estimates is calculated from the known observability, instrumentation quality, and number of tests. Prior information and tests from development testing (DT) could also be used in the beginning until an adequate number of post deployment tests could be accumulated. Periodic reassessment of the test program adequacy to estimate the MOPs and associated confidences may require feedback to the planning stages to reassess the confidence requirements.

Next, the system evaluator predicts the MOE and confidence bounds for the required reference set of scenarios, using the force level simulations to flow up the MOPs (and confidences bounds) to MOE (and confidence bounds). He conducts model fault isolation to determine which MOP is out of specification and its resultant contribution to the MOE. Periodic reassessment of the test program adequacy for current MOE requirements must be done.

Finally, the system evaluator conducts Force Level Evaluations with latest estimated models by using force level simulations to flow up the estimated MOPs (and confidences bounds) to MOE (and confidence bounds) to evaluate the adequacy of the systems for many different campaigns. This allows trade offs to be made for optimum planning of the BMD FoS deployment, as is illustrated in Mitchell et.al. (2001). He also develops and updates a functionalized performance prediction model to be used in the real-time employment of the BMD response to an operational threat.

A number of "grand" technical challenges must be faced in order to optimally build in confidence to the BMD FoS. The specter of limited testing will force heavy reliance on more "physics based" models to optimally extract the maximum information from each test. System, subsystem, and potentially lower level testing will have to be combined. Reliability modeling may have to be revised. Methodologies for semi-automatically optimizing the test programs design and optimally combining the diverse types of testing will be needed.

<u>Model Estimation vs. Model Validation</u>

Note that this process provides an "estimated" model from the test data, which is distinct from a "validated" model. Model validation is focused on how well the model predicts the real world over a limited set of test points (hopefully, but not usually, the intended use of the system). A hypothesis test is conducted on the test data to invalidate the model. Confidence in the model at the test points is not transferable to non-tested conditions. Model estimation directly estimates the model parameters from the test data. A "physics based" model (with scenario independent parameters) is fit directly to all relevant test data (system, subsystem, etc.) for <u>statistical leveraging</u>. The computed error statistics of the scenario independent model parameters indicate which part of model is poorly known and allows tradeoffs in instrumentation type, quality, and test sizing to meet system evaluation requirements. Most importantly, the parameter estimates and the error statistics can be transferred to any <u>non-testable</u> MOE estimate (via engagement scenario simulations) for confidence bound predictions.

Conclusions

The effectiveness of our deployed BMD FoS needs to be assured. <u>Confidence based model building from T&E</u> is the key for <u>credibly</u> predicting performance. The process starts with proper performance <u>evaluation</u> requirements (confidence) to define the test program and analysis methodologies. Estimated models, with confidence, are developed from testing. These models, with confidence, are extrapolated to operational, untested conditions with confidence. This allows for optimum planning for placement and usage of assets before the action commences as well as optimum real-time threat response.

References

IDA/WSEG Report 92A (1966). *Guidelines for Evaluating Operational Test Programs (U)*, May 1966, SECRET

Levy, L.J. (1992*). Large-Scale System Performance Prediction with Confidence from Limited Field Testing Using Parameter Identification*," with D. W. Porter, *Johns Hopkins APL Technical Digest*, Vol. 13, No. 2, pp. 300-308.

Mitchell, B.L., S.E. Brown, and T.P. Spriesterbach (2001). *A Methodology to Extend BMDO Test Results Using Force-on-Force Simulation*, Presented at the 2nd Biennial National Forum on Weapon System Effectiveness.

Simkins, L.S., G.J. Quart (1990), *Trident II Accuracy Evaluation Overview (U)*, APL Technical Review, vol. 2, # 2, SECRET.

US Strategic Command Directive (1998), *Guidelines for Nuclear Weapon System Operational Testing and Reporting (U*), 23 March 1998, Confidential.

Larry Levy is Chief Scientist of the Strategic Systems Department of the Johns Hopkins University Applied Physics Laboratory. He received a Ph.D. in Electrical Engineering from Iowa State University in 1971. He has worked on applied Kalman filtering and system identification for over 30 years and was the co-developer of SATRACK (a GPS based missile tracking instrumentation system) and the methodology for Trident II accuracy evaluation with confidence. E-mail: larry.levy@jhuapl.edu